

J. Clin. Chem. Clin. Biochem.

Vol. 27, 1989, pp. 433–443

© 1989 Walter de Gruyter & Co.
Berlin · New York

On the Performance and Reliability of Mechanized Urine Teststrip Measurement in Comparison with Visual Reading

By A. M. Marx, J. Kropf and A. M. Gressner

Abteilung für Klinische Chemie und Zentrallaboratorium, Klinikum der Philipps Universität Marburg, Marburg

(Received January 3/April 25, 1989)

Summary: Prescreening of urine specimens by teststrips is a valuable procedure for reducing the work load of the urine analysis laboratory: positive results for leukocytes, erythrocytes (haemoglobin), protein, and/or nitrite are widely used to select pathological specimens for subsequent microscopic examination. By standardization of the measurement conditions, mechanized teststrip reading is claimed to give more reproducible results than conventional techniques. To assess their ability to improve urine prescreening, especially with regard to the comparability of the results, the practical and analytical performance of three commercially available analysers (Rapimat® II/T from Behringwerke AG, Urotron® RL9 from Boehringer Mannheim GmbH, and Clinitek® 200 from Ames/Bayer Diagnostic) was compared with visual reading. Analytical criteria were assessed using routine urine samples, while reproducibility was tested by repeated analysis of three different commercial control urines (Kova®Trol from Madaus). A mean imprecision between 3% and 11.9% was found for the mechanized dipstick reading which was comparable to that found with visual examination (4.5% with Combur⁹-teststrips, Boehringer Mannheim GmbH). Due to the crude classification of the results, the different analysers as well as the visual technique gave quite different distributions for each of the semiquantitative parameters in the same urine samples. Even if statistical analysis was restricted to the frequency of positive results only, significant differences (χ^2 -test, $p < 0.001$ – 0.05) between methods were obtained, but these differences could not be attributed to one method alone. Using microscopic sediment analysis as reference, pathological urines were detected with a comparable sensitivity/specificity: Urotron 0.85/0.84, Rapimat 0.91/0.67, Clinitek 0.82/0.81, and duplicate visual reading 0.88/0.67 and 0.91/0.93. Mechanized teststrip reading had no obvious advantage with respect to the time required. We conclude:

- (i) no improvement in analytical performance or in speed of analysis could be claimed for mechanized methods in comparison with visual reading;
- (ii) mechanized teststrip reading might decrease the work load of the urine laboratory if integrated into a computerized laboratory system;
- (iii) mechanized teststrip reading will become analytically advantageous over visual reading if a more refined classification of the results is achieved.

Introduction

Pre-screening of urine specimens by teststrips is a valuable tool for reducing the work load in the urine analysis laboratory (1, 2). Positive results for leukocytes, erythrocytes, protein and/or nitrite are widely used to select pathological samples for microscopic examination (3–6). In the last years various devices

for mechanized teststrip reading have been developed which are considered to have several advantages in relation to visual inspection. For example, they are claimed to give more reproducible results by standardized incubation and measurement conditions. Including multi-centre evaluation, the reliability of the common instruments and their precursors has been

estimated by several authors (7–12). They found a good correlation between the "semiquantitative" results of the tested devices and values obtained by quantitative reference methods. The high validity of the results thus documented for mechanized dipstick reading, should be followed by a comparison of results from different laboratories using instruments from different manufacturers. We therefore investigated whether mechanized teststrip reading itself is able to improve urine prescreening, particularly with respect to the comparability of the analytical results. Since mechanized dipstick reading is also expected to increase the rate of analyses, some aspects of the routine performance were also considered.

In an initial trial we assessed three different mechanized analyses and a visual teststrip method, and compared them with those of our routinely applied visual teststrip analysis. In a second trial the same samples were consecutively analysed by all methods and the results correlated with those of the corresponding sediment analysis.

Materials and Methods

Urine samples

Unselected urine specimens presented for routine analysis to the Central Laboratory of the University Hospital Marburg were used for the investigation. They were completely processed within one hour after receipt in the laboratory. Reproducibility of the methods was tested using three different control urines (Kova®Trol I–III) purchased from Madaus Diagnostik, Cologne, FRG.

Urine prescreening

The results of the following methods for urine prescreening were compared:

- (i) Routine urine analysis, which was performed by visual examination of the Combur⁹-Teststrip from Boehringer Mannheim GmbH, Mannheim, FRG.
- (ii) Second manual urine analysis by an independent technician, also using Combur⁹-Teststrip.
- (iii) Mechanized examination of the Rapignost® Total-Screen L test-strips by the Behring Rapimat II/T supplied by Behring Werke AG Marburg, FRG.
- (iv) Mechanized examination of the test-strip Combur⁹-Test® RL by the Urotron® RL9 obtained from Boehringer Mannheim GmbH, Mannheim, FRG.
- (v) Mechanized examination of the MultistixTM 10 SG by the Clinitek® 200 from Bayer Diagnostic and Electronic, Munich, FRG.

The analytical procedure of the tested instruments for mechanized teststrip reading is based on reflectance photometry of teststrips. After application, the teststrip is transported to a detection unit, either by a moving rake (Clinitek) or by adherence to a moving transport paperstrip (Urotron, Rapimat). Illuminated by LEDs the coloured testfields reflect light in relation to the concentration of the respective analyte. On the

basis of a reference curve for each analyte, this reflectance, measured by a photodetector, is converted into a concentration range. The colour intensity of the test fields depends on the incubation time, which is standardized by the fixed working cycle of the analytical procedure. The sluggish colour development of the leukocyte testfield is a well known phenomenon (10), and an attempt is made to compensate for this by prolonged incubation period (Clinitek) or by selective heating of this field (Rapimat).

Any positive teststrip reading for leukocytes, blood, protein, and/or nitrite was considered as pathological.

Urine specimens presented for routine analysis (about 12 ml) were used in our study. In the second trial, however, we inserted five different teststrips into each urine. This procedure might be expected to cause a significant sample contamination (13). To study a possible effect on analysis, we incubated teststrips for up to 10 min in fractionated urines (normal and pathological). This caused no noticeable difference in subsequent teststrip results, which showed that cross contamination between test strips is not analytically significant.

Urine sediment analysis

Microscopic sediment examination was carried out according to routine laboratory procedures (14): 10 ml urine were centrifuged for 5 minutes at 2000 min⁻¹. After aspiration of the supernatant, one drop of the resuspended urine sediment was examined by microscopy. The urine sediments were classified as positive (pathological) in two different ways:

- (a) standard sediment: more than two erythrocytes (visual field, 400x), more than five leukocytes (visual field, 400x), pathological casts, and/or significant bacteriuria.
- (b) extended sediment: on the supposition that the agreement between at least four independent methods minimizes the chance of a false positive result, sediments were defined to be positive if:
 - positive teststrip results for leukocytes and/or erythrocytes were found by all five prescreening methods.
 - positive teststrip results for leukocytes and/or erythrocytes were found by at least four methods in combination with respective traces found by microscopic examination.

Statistics

For statistical calculations and documentation of the results the software packages SAS (SAS Institute Inc, Cary, USA; 15), Statgraphics (STSC Inc., Maryland, USA) and Freelance (Lotus Development Corp., Windsor, UK) were utilized.

Results

First trial

The teststrip results are supposed to be an objective and reliable measure for the content of the tested analytes in the urine sample. This should imply a high degree of coincidence between the obtained results, independent of the type of teststrip or the procedure used for its evaluation. In an initial trial we therefore performed analyses on three different instruments for mechanized teststrip examination, and compared the results with those of our routinely applied visual teststrip analysis.

Independent series of urines were used for the evaluation. The results were cross-tabulated with respect to qualitative levels of concentration, quoted for the different teststrips by the manufacturer. As an example, figure 1 shows the results of the analysis of erythrocyte-content obtained by the Rapimat (a) and the Urotron (b) in comparison with those of the corresponding routine-analysis. The cases of coincidence are given in the main-diagonals falling from left to the right (undotted fields). The fraction of coincidence was calculated from the number of cases on this diagonal and the total number of tests. For all analytes, this fraction of coincidence with routine analysis is given in table 1. Coincidence (mean 0.79) differed for the single analytes, being low for pH (0.44–0.52) and highest (nitrite: 0.85–0.97) for analytes with binary classification. The frequency of total agreement across the combined analytes was found to be very low (0.32–0.56).

Results that differ from those of our routine analysis might be explained by random fluctuations or by the different sensitivity of one method. In the first case a symmetrical distribution of the results around the diagonal (slightly dotted) is to be expected (fig. 1a). Cases of gross discrepancies between the compared methods are placed in the darkly dotted fields. In the case of different sensitivities, most of the results should be found either above or below the main diagonal (fig. 1b). In table 1 the percentages of such over- or underestimations with respect to the results of our routine analysis are given in brackets.

Second trial

During the first trial several discrepancies between the results of the routine urine analysis and of the evaluated methods were observed. We therefore carried out a second trial, during which the same 159 unselected routine urine samples were analysed by all the three instruments and visually by the routine staff and a second examiner. For this trial two of the instruments had to be exchanged: the Rapimat because of an instrumental failure and the Clinitek for

Routine visual reading

a

Rapimat

	-	+	++	+++
-	137	16	1	
+	4	19	1	
++		6	2	
+++		5	7	5

Routine visual reading

b

Urotron

	-	+	++	+++
-	91			
+	8	17		
++	1	23	1	
+++		5	7	3

Fig. 1. Results of the analysis of erythrocyte content obtained by the Rapimat (a) or the Urotron (b) compared with the results of the corresponding routine analysis. The frequencies of the respective result combinations are cross-tabulated.

reasons of organization. To evaluate the validity of the teststrip results, microscopic sediment examination was carried out for all samples.

Coincidence in concentration results

The results of pH-analysis are given with different numbers of subclasses as defined by the respective manufacturer. Even if this fact is considered, quite different distributions of pH-results were obtained for the same group of urines (fig. 2). For the other analytes also, quite different concentration ranges are specified by the respective teststrip manufacturers.

Tab. 1. *First trial*: Fractions of coincidence of the results obtained by mechanized teststrip reading compared with the results of the visual routine analysis. The numbers in brackets indicate the percentage of clear over- (+) or under- (−) estimations. Given are the fractions for the single analytes, the mean fraction of coincidence, and the fraction of samples with coincidence in all analytes (All).

	Fractions of coincidence with routine analysis								Mean	All
	Leuko-cytes	Nitrite	pH	Protein	Glucose	Urobilinogen	Bilirubin	Erythrocytes		
Clinitek	0.82 (+16)	0.95	0.44	0.60 (+38)	0.91	0.88 (+10)	0.93	0.76 (−22)	0.79	0.32
Urotron	0.75 (+18)	0.97	0.52	0.80 (−17)	0.86	0.81 (+17)	0.91	0.72 (+28)	0.79	0.56
Rapimat	0.82 (−14)	0.85 (+15)	0.47	0.84	0.90	0.80	0.92	0.80	0.80	0.38

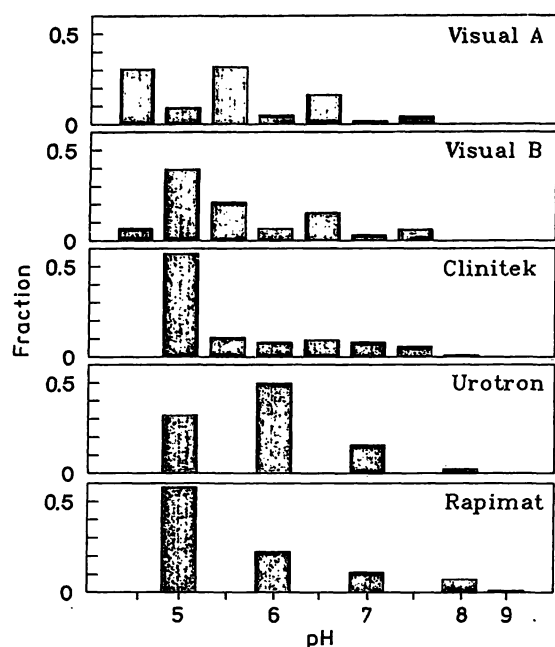
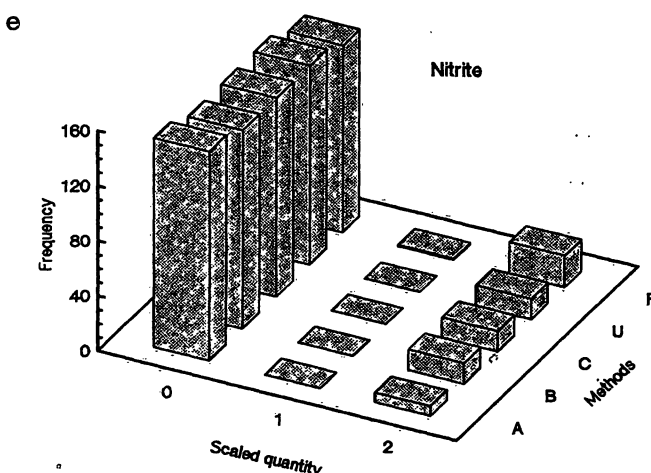
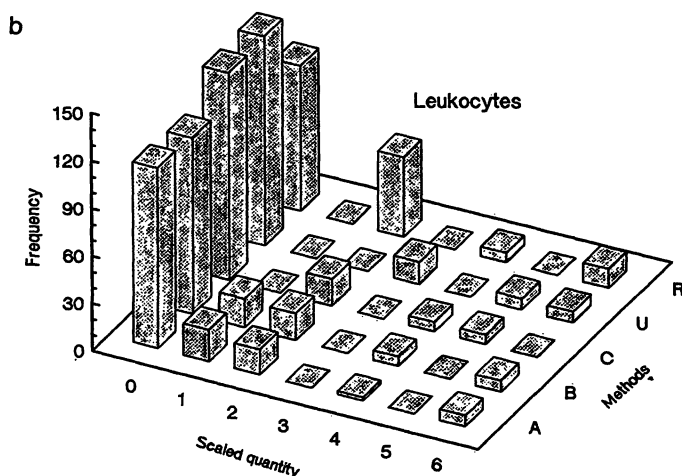
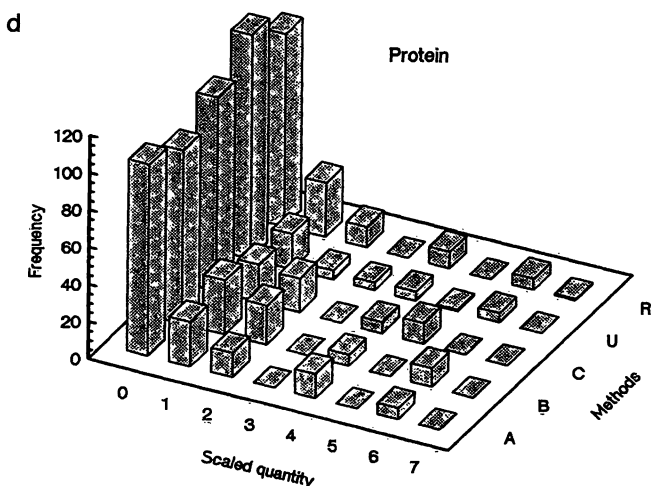
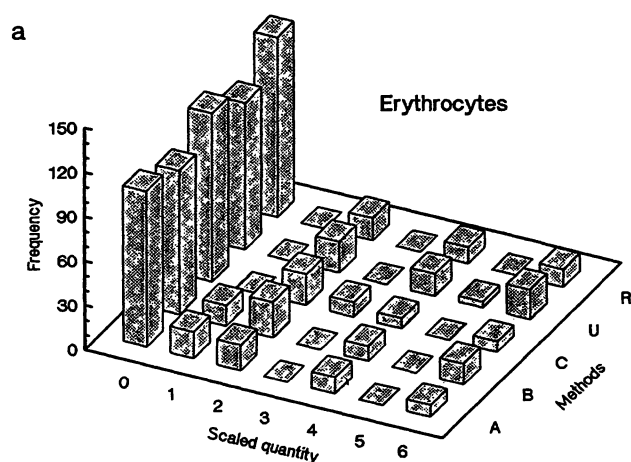
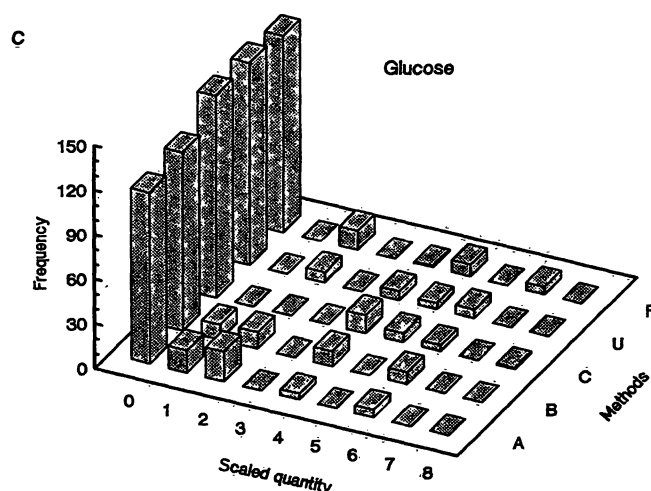


Fig. 2. Distribution of pH results measured within the same group of urines ($n = 159$). The fractions of the respective pH results are given.

This eventually could be a reason for the discrepancies among the results observed during the first trial. To enable an unbiased comparison of the measured values, we therefore scaled the concentration ranges according to table 2. All the teststrip results obtained were converted into these units. In figure 3, the frequency distributions of these units are shown, as detected by the different methods within the tested group of urines.



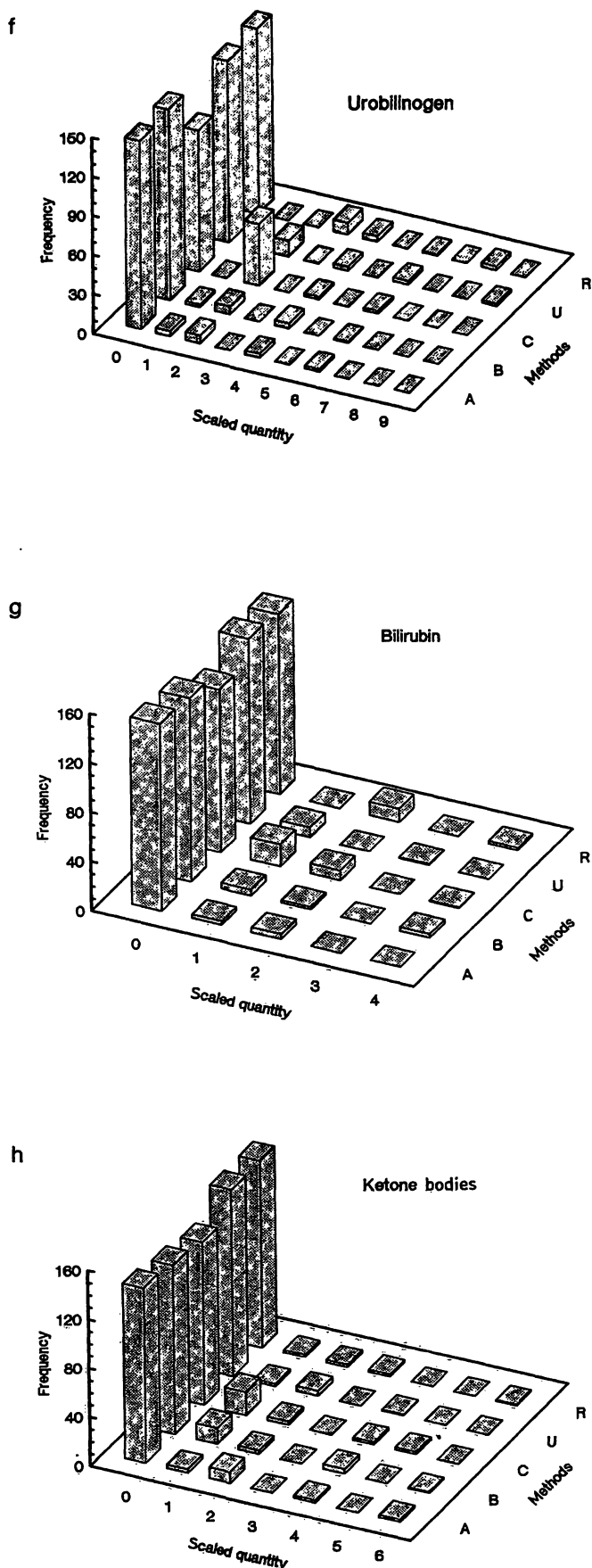


Fig. 3. Frequency distributions of uniformly scaled teststrip results found within the same group ($n = 159$) of urines. A: routine visual reading, B: second visual reading, C: Clinitek, U: Urotron, R: Rapimat.

Tab. 2. Attribution of the individual cutoff values given for all analytes by the different manufacturers (converted to SI-units) to a uniform scale defined by the authors.

Analyte	Cutoff value specified for the teststrips used by				
	Clinitek	Rapimat	Urotron	Visual	Scale
Leukocytes ($10^6/l$)	Neg	Neg	Neg	Neg	0
	—	—	—	Trace	1
	15	20	25	+	2
	70	75	—	++	4
	125	—	100	—	5
	500	500	500	+++	6
Nitrite ($\mu\text{mol/l}$)	Neg	Neg	Neg	Neg	0
	13	11	11	11	2
Protein (g/l)	Neg	Neg	Neg	Neg	0
	0.15	0.15	0.15	Trace	1
	0.30	0.30	0.30	+	2
	—	—	0.60	—	3
	1.00	1.00	1.00	++	4
	3.00	—	3.00	—	5
	20.00	5.00	5.00	+++	6
Glucose (mmol/l)	Neg	Neg	Neg	Neg	0
	—	—	—	Trace	1
	—	2.8	2.8	+	2
	5.55	—	5.55	++	4
	—	8.33	11.10	—	5
	13.88	—	16.65	+++	6
	27.75	27.75	—	—	7
	55.50	—	—	++++	8
	111.0	—	—	—	9
Ketone bodies (g/l)	Neg	Neg	Neg	Neg	0
	0.05	—	—	Trace	1
	0.15	0.10	0.10	+	2
	—	0.25	—	—	3
	0.40	—	0.50	++	4
	0.80	—	—	—	5
	—	1.00	—	+++	6
	1.60	—	1.50	—	7
	—	3.00	—	—	8
Urobilinogen ($\mu\text{mol/l}$)	Neg	Neg	Neg	Neg	0
	1.8	—	—	Trace	1
	18	—	18	+	2
	35	35	—	—	3
	70	70	70	++	4
	140	140	140	+++	5
	210	210	—	++++	6
	—	—	280	—	8
	—	—	—	—	9
Bilirubin ($\mu\text{mol/l}$)	Neg	Neg	Neg	Neg	0
	—	—	8.6	Trace	1
	slight	17	—	+	2
	—	—	26	—	3
	medium	34	—	++	4
	—	—	51	—	5
	strong	68	—	+++	6
	—	—	103	—	7
Erythrocytes ($10^6/l$)	—	—	205	—	8
	Neg	Neg	Neg	Neg	0
	—	—	—	Trace	1
	10	10	10	+	2
	25	—	—	—	3
	80	60	50	++	4
	—	—	150	—	5
	200	300	250	+++	6

It is quite obvious that, even after conversion of the results into an uniform scale, for the same group of urines, differences in the distribution of concentrations were obtained. Particularly striking is the high proportion of positive leukocytes, nitrite and urobilinogen results found by the Rapimat or Clinitek, respectively. The percentage of coincidence and the frequencies of clear over- and underestimations with respect to the results of routine analysis were calculated as described for the first trial and are given in table 3. Most results of the first trial were confirmed by the second trial. In both trials the Urotron was less sensitive in the detection of protein and more sensitive in the detection of erythrocytes and urobilinogen. The Clinitek and the Rapimat proved to be more sensitive in the detection of urobilinogen and of nitrite, respectively. The higher detection rate of protein and the lower detection rate of erythrocytes by the Clinitek was not confirmed in this trial. Unexpected were the results for leukocytes: while in the first trial the Clinitek and the Urotron were more, and the Rapimat was less sensitive than our routine analysis, in the second trial the results were reversed. Some of these differences between the results of the first and the second trial might be attributed to the fact that two of the tested instruments had been exchanged (see above).

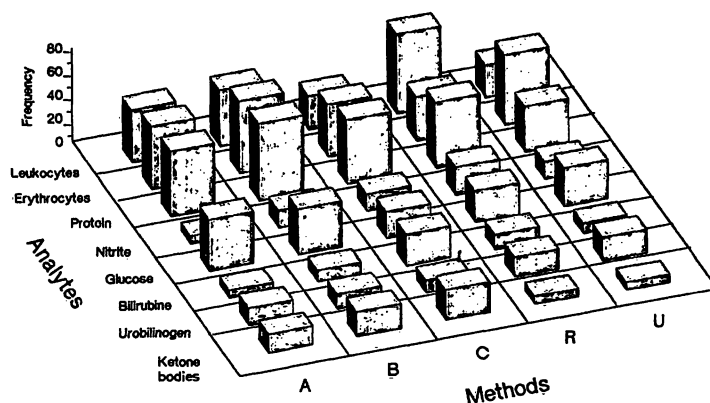


Fig. 4. Frequencies of positive teststrip results found by the compared methods within a group of 159 urines. A: routine visual reading, B: second visual reading, C: Clinitek, U: Urotron, R: Rapimat.

We even found gross differences between the evaluated methods for the frequencies of positive results (fig. 4). As an example, for the same group of 159 urines, positive leukocytes teststrip results were found in 45 and 48 urines by visual reading, in 28 or 29 urines by the Urotron and Clinitek, respectively and in 68 urines by the Rapimat. Similar discrepancies were found for nearly all analytes. The statistical analysis (χ^2 -test, $p < 0.05$) of these differences is given for several comparisons in table 4. Although there are

Tab. 3. *Second trial*: Fractions of coincidence of the results obtained by mechanized teststrip reading compared with the results of the visual routine analysis. The numbers in brackets indicate the percentage of clear over- (+) or under- (-) estimations. Given are the fractions for the single analytes and the mean fraction of coincidence.

	Fractions of coincidence with routine analysis							Mean
	Leuko- cytes	Nitrite	Protein	Glucose	Urobili- nogen	Bili- rubin	Erythro- cytes	
Visual ^a	0.71	0.95	0.69	0.77	0.89	0.93	0.72 (+20)	0.81
Clinitek	0.84 (-13)	0.95	0.74	0.80	0.74 (+25)	0.84	0.83	0.82
Urotron	0.83 (-13)	0.95	0.74 (-24)	0.79	0.87 (+10)	0.93	0.75 (+23)	0.84
Rapimat	0.74 (+24)	0.89 (+11)	0.72	0.82	0.88	0.91	0.80	0.82

^a independent second visual reading

Tab. 4. Results of χ^2 -test for differences in the frequencies of positive teststrip results obtained with respect to several analytes. A: routine visual reading, B: second visual reading, C: Clinitek, U: Urotron, R: Rapimat. Comparisons: First column: all methods. Second column: devices versus visual reading. Third column: devices only. Fourth column: visual reading only. Fifth column: Combur 9 teststrip versus the other two teststrips. Sixth column: Combur 9 teststrip only. Seventh column: Rapignost versus Multistix. Significances: n. s. not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Analytes	U/A/B/C/R	UCR/AB	U/C/R	A/B	UAB/CR	U/A/B	C/R
Leukocytes	***	n. s.	***	n. s.	n. s.	*	***
Nitrite	*	*	n. s.	n. s.	*	n. s.	n. s.
Protein	*	*	n. s.	n. s.	n. s.	**	n. s.
Glucose	**	***	n. s.	n. s.	*	**	n. s.
Ketone bodies	***	n. s.	***	n. s.	n. s.	*	***
Urobilinogen	***	***	***	n. s.	***	n. s.	***
Bilirubin	***	*	***	n. s.	***	n. s.	**
Erythrocytes	*	n. s.	*	n. s.	**	n. s.	n. s.

significant differences for all analytes within the group of all the evaluated methods, surprisingly, there is not even one analyte, for which the frequency of positive results significantly differs between the two independent visual readings. For erythrocytes there are no significant differences in the frequency of positive results between the Urotron and the two visual readings or between the Clinitek and the Rapimat. But the differences are significant, if the Urotron and both visual readings are tested versus the Clinitek and Rapimat. Since the Urotron and visual reading use the same kind of teststrip, these differences might be attributed to a different sensitivity of the Combur 9 teststrip. Similar results are obtained for the analysis of nitrite. Although the same kind of teststrip is used, there is a significant difference in the frequency of positive protein results between the Urotron and the two visual methods. Since the results of visual reading are comparable to those of the Clinitek and Rapimat (fig. 4), the mechanized teststrip reading by the Urotron gives a remarkably lower number of positive protein results.

Comparison of agreement with sediment findings

Because of the gross discrepancies in the frequencies of positive teststrip results (see above), we compared the teststrip results for erythrocytes and leukocytes with the findings in the corresponding sediment. Sediments were considered to be positive, if more than two erythrocytes or more than five leukocytes were found in the visual field (400x). The results are shown in table 5. In the detection of leukocytes the evaluated methods can be divided into three groups:

- (i) the Rapimat was significantly (χ^2 -test, $p < 0.001$) more sensitive with only 2% of false negative results. But in contrast to the observation of Kutter (10), we noticed a high fraction (0.21) of false positive results;
- (ii) the Clinitek and Urotron were less sensitive, showing 0.10 and 0.12 false negative leukocyte results, respectively; however, these instruments also gave the lowest fractions of false positive results (0.04 and 0.06);
- (iii) visual reading lies between the former two groups with respect to the fractions of false positive and negative results.

Because of the high portion of false positive results, the significantly (χ^2 -test, $p < 0.001$) lowest analytical efficiency of 0.77 was found for the Rapimat. For erythrocytes there were no significant differences with respect to analytical efficiency. If all methods are compared, there are significant (χ^2 -test, $p < 0.05$) dif-

Tab. 5. Validity of erythrocyte/leukocyte detection by teststrip analysis when referred to sediment analysis. A: routine visual reading, B: second visual reading, C: Clinitek, U: Urotron, R: Rapimat. Significances: *** $p < 0.001$ ($n = 159$).

Leukocytes Frequency of	Methods				
	A	B	C	U	R
True positive	30	30	22	19	35
True negative	106	103	114	112	88
False positive	15	18	7	9	33
False negative	8	8	16	19	3***
Efficiency	0.86	0.84	0.86	0.82	0.77***

Erythrocytes Frequency of	Methods				
	A	B	C	U	R
True positive	40	40	34	41	30
True negative	99	91	102	94	105
False positive	14	22	11	19	8
False negative	6	6	12	5	16
Efficiency	0.87	0.82	0.86	0.85	0.85

ferences in the frequency of false positive erythrocytes results. But these differences cannot be attributed to one method alone. Nevertheless, the Rapimat and Clinitek seemed to be somewhat less sensitive than the other methods. We were not able to judge the influence of ascorbic acid (a potent inhibitor of the erythrocytes test field (10)) on erythrocyte detection: the number of urines with a significant (> 1.1 mmol/l) ascorbic acid content was too low for an objective evaluation.

Comparison of urine prescreening results

Positive teststrip results for erythrocytes, leukocytes, nitrite and/or protein are widely used to select pathological urines for sediment analysis. We therefore evaluated the applicability of the different teststrip results for urine prescreening. Pathological teststrip results and positive sediment findings were defined as explained in Materials and Methods. Pathological teststrip results were classified as true positive if verified by a corresponding positive sediment finding. The results and some statistical measures are shown in tables 6a, b. All evaluated methods show a remarkably high fraction (mean: 0.21) of false positive teststrip results and corresponding low diagnostic specificities (mean: 0.64, tab. 6a). This can be explained in part by lysis of cellular components leading to positive teststrip results without any leukocytes and/or erythrocytes in the sediment. Since cell lysis has also been considered by others (5), we therefore extended the

Tab. 6. Validity of the prediction of pathological sediment findings by teststrip analysis.

a: Reference standard sediment,
b: Reference extended sediment (see Material and Methods).

A: routine visual reading, B: second visual reading,
C: Clinitek, U: Urotron, R: Rapimat.

a Frequency of:	Methods				
	A	B	C	R	U
True positive	62	62	56	62	59
True negative	67	52	61	49	63
False positive	24	39	30	42	28
False negative	6	6	12	6	9
Sensitivity	0.91	0.91	0.82	0.91	0.87
Specificity	0.74	0.57	0.67	0.54	0.60
Efficiency	0.81	0.72	0.74	0.70	0.76
Pred. value +	0.72	0.61	0.65	0.60	0.68
Pred. value -	0.92	0.90	0.84	0.89	0.88

b Frequency of:	Methods				
	A	B	C	R	U
True positive	81	78	73	81	76
True negative	65	47	57	47	59
False positive	5	23	13	23	11
False negative	8	11	16	8	13
Sensitivity	0.91	0.88	0.82	0.91	0.85
Specificity	0.93	0.67	0.81	0.67	0.84
Efficiency	0.92	0.79	0.82	0.81	0.85
Pred. value +	0.94	0.77	0.85	0.78	0.87
Pred. value -	0.89	0.81	0.78	0.85	0.82

criteria for a positive sediment, taking partly into account the results of the teststrip reading itself. Beside cell lysis this procedure corrects for potential mistakes in the evaluation of urine sediment. With no noticeable effect on sensitivity, this modification reduced the fraction of false positive results (tab. 6b), i.e. 0.09 of the results proved to be false positive, leading to a mean specificity of 0.78. Although, with respect to false positive results, the differences between the evaluated methods were significant (χ^2 -test, $p < 0.002$), these differences could not be attributed to one method alone; routine visual reading and the Urotron produced a remarkably low (5 and 11 respectively) number of false positive results. In contrast, the second visual reading procedure and the Rapimat showed a high frequency of false positive results. Without significant differences, a mean fraction of 0.07 (Clinitek: 0.10, Rapimat and routine analysis: 0.05) of the positive sediment findings remained undetected by the teststrip prescreening. There were significant differences (χ^2 -test, $p < 0.02$) between the evaluated methods with respect to analytical efficiency. The highest efficiency (0.92) was found for our routine analysis, while the efficiencies of the other methods were not significantly different, with a mean

value of 0.82. Since the second visual reading had the lowest efficiency of 0.79 there were no significant differences in efficiency if visual teststrip examination was compared with mechanized urine prescreening. Within the mechanized methods the Rapimat achieves highest sensitivity (0.91), but lowest specificity (0.67) and efficiency (0.81). Best performance with respect to all criteria was given by the routine visual method.

Reproducibility

Reproducibility (imprecision) should be tested by the use of suitable control materials. Special control urines for teststrip analysis, which are claimed to ensure well defined concentration ranges for a long period, are commercially available. To test the imprecision of the evaluated methods, we serially analysed three different control urines ten times (normal, slightly pathological, and pathological). To avoid recognition effects in visual reading, the 30 control urine samples were interspersed randomly and unknown to the examiner with the routine urine samples. The concentration ranges given for the respective analytes by the control urine manufacturer are very large (e.g. bilirubin in Kova®Trol I: 25–137 $\mu\text{mol/l}$, arbitrary ranges between 1+ and 3+ for different instruments). For this reason, figure 5 shows only the percentages of results deviating from the expected values (most frequently obtained by the respective method) for each analyte. With some exceptions, imprecisions of the evaluated methods are comparable (mean 4.1%, 3.0% and 4.5% of the Rapimat, Clinitek, and visual reading, respectively). The twofold higher imprecision found for pH-results of the Clinitek might be due to the doubled number of pH-subclasses given by this

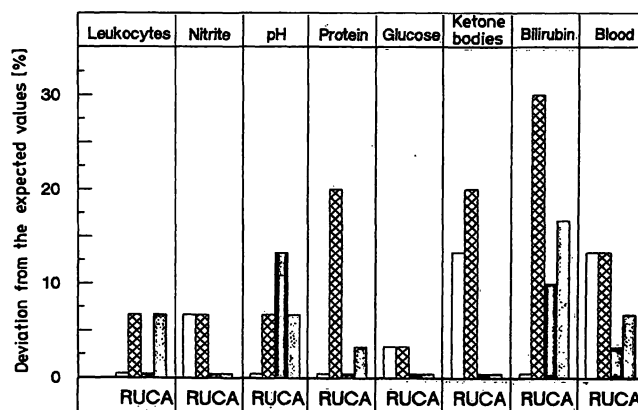


Fig. 5. Imprecision in the analysis of three different control urines. The percentages of deviation from the expected values are given (i.e. the value most frequently obtained by the respective method).
R: Rapimat, U: Urotron, C: Clinitek, A: routine visual reading.

device. Unexpected high imprecision for protein, ketone bodies (20%) and bilirubin (30%) was found for the Urotron, leading to a mean of 11.9%. At least in part, this might be explained by concentration values of these analytes near the respective decision limits (6).

Speed of analysis

The time used for the examination of a urine teststrip by mechanized instruments is defined by their working cycle (Rapimat and Clinitek: 10 s, Urotron: 12 s). We measured the time required for visual reading by a well trained technician. The results of 50 measurements (selected by chance) are given in table 7. Visual reading time (between 2 and 12 s, mean 5 s) was influenced by the count of positive test fields/teststrips and/or the count of results near a cutoff point.

Tab. 7. Time required for the visual examination of urine teststrips by a well trained technician. The time was measured from the end of the incubation period until the end of the result documentation. Given are the frequencies found for different time intervals within an arbitrary chosen group of 50 routine urines.

Visual teststrip reading time					
Time required	1–2 s	3–4 s	5–6 s	7–8 s	9–12 s
Frequency	13	7	12	16	2

Assessment of practicability

During our study several observations on the practicability and reliability of the three instruments were made. Generally, mechanized teststrip reading has one disadvantage: after application of the teststrip to the instrument, it mostly will not be further controlled by the user. If the colour development of a testfield is disturbed for example by interfering substances, this is not observed by the user or detected by the instrument. Only the Clinitek offers the facility for observing the teststrip during the whole analysis. But we noticed that, particularly under a heavy workload, further monitoring of the teststrip is often omitted by the user. Other observations are specific for the individual instrument:

Clinitek

In the view of its practicability this device was clearly favoured by all participants. The instrument is simply operated by means of only four keys and enables easy handling during teststrip application. In addition, the fact that the prompt for insertion of the teststrip into the urine is given by an audible signal in combination with an optical countdown was considered to be very

convenient, especially in the presence of high background noise. Although the daily cleaning procedure requires removal of some parts of the instrument it was judged to be most reliable. The Clinitek has to be calibrated daily with reusable calibration strips. Individual sample identification was judged to be very time consuming. The reprint function of the delivered apparatus showed a malfunction and thus could not be evaluated.

Rapimat

This device offers some additional features. For this reason, if compared with the Clinitek, handling of the instrument is more complicated. For example, the reprint of stored results is only possible after calling the routine for input of microscopic results. Teststrip application was judged to be somewhat cumbersome: the strips have to be placed exactly flushed in a defined position on the transport paper during a fixed time interval (5 seconds). Most users found fault with the fact that the applied teststrips cannot be observed after being transported into the instrument. A clear advantage provided only by the Rapimat is the highlighting of pathological results by red coloured printing. The manufacturer recommends the daily cleaning of the teststrip transport system which mainly means the simple replacement of the transport paper. The sensor head of the measurement system has to be removed for cleaning once a week. We believe that this more complicated procedure should also be done daily, especially under heavy work load. During evaluation the initially delivered device had to be replaced because of malfunction of the teststrip detection unit. Some Rapignost teststrips of the evaluated lot (633241A) showed coloured inclusions within some testfields, but we did not notice any negative influences on the results.

Urotron

Design and functional attributes of this device are similar to those of the Rapimat. Thus most of the comments on the Rapimat also apply to the Urotron, with the following exceptions: reprinting of stored results is a little easier and the timing of teststrip application is less stringent. Compared to the other teststrips, the Combur⁹ RL requires larger urine volumes and longer sample tubes:

- (a) the single testfields extend over a greater distance of the stick, and
- (b) the rigid supporting layer makes it more difficult to adapt the strip to small volumes by slight folding.

A further observation was a slight bending of the wet teststrip. Sometimes this bending caused a malfunction of the teststrip transport. Especially in such cases the impossibility of visually monitoring the sticks during measurement was considered by all users to be a disadvantage.

Discussion

Introduction of a new method or a new instrument for analysis is recommended only if it permits analyses to be performed more accurately and precisely, more sensitively, at a higher rate of throughput or more cheaply, or if it is more economical of manpower. In the field of urine prescreening, instruments for mechanized teststrip reading have been developed, which are considered to have several advantages in relation to visual inspection, e.g. to give more reliable and reproducible results and a higher speed of analysis. Although the reliability of these instruments individually has been evaluated by several authors (7–13), up to now a comparative study on the advantages of mechanized reading versus visual teststrip examination is lacking. We therefore studied three commercially available instruments for their ability to improve urine prescreening.

While it is generally believed that visual judgment of teststrips has a low precision, this was not verified in the present study. As shown in figure 5, the reproducibility of mechanized reading is not significantly enhanced over visual teststrip evaluation. This might be explained by the fact that the results of teststrip reading are classified into a few concentration ranges. Small variations caused by irreproducibility of the colour development of the test fields and/or of the teststrip reading will usually be covered by this rough classification, giving an acceptable imprecision of about 4%. But due to this classification, reproducibility will be strongly influenced by sample concentrations near a decision point; in the worst cases, reproducibility can be as low as 50% for all kinds of teststrip reading (12). This is probably a reason for the decreased reproducibility of the Urotron which we found for the analysis of protein, ketone bodies, and bilirubin. On the other hand *Franzini et al.* (9) also reported a coefficient of variation between 10% and 19% for the Urotron. But they calculated this value directly from the measured reflectance values and it is possible that this kind of calculation would also increase the imprecision of the other instruments. A further explanation for decreased precision might be instrumental malfunction caused, for example, by contamination of the detection unit. This possibility stresses the necessity of a regular quality control in mechanized teststrip reading.

With respect to urine prescreening the detection of pathological urines should be possible in an objective and comparable manner and should be independent of the applied method. Thus, with respect to positive results, agreement is to be expected between all methods, including mechanized teststrip reading. Contrary to this expectation we found significant differences in the frequencies of positive results between the evaluated methods, for the same group of urines (fig. 4, tab. 4). We did not test in detail whether these different results have to be attributed to differences in the sensitivity of the teststrip itself and/or its reading. However, the fact that significant differences were found clearly stresses the necessity of a suitable control material. We believe that such a material is not yet available, since to our knowledge all the commercial control urines have concentration ranges that are too large (see above). Thus, unknown to the physician, at least some diagnostic decisions might be influenced by the kind of teststrip used and/or the reliability of its examination.

Two of the instruments for mechanized teststrip reading offer the facility for adapting the cutoff values of the reference curves, but enhancement of sensitivity by means of cutoff value adjustment consequently lowers specificity. The Rapimat gave the significantly lowest number of false negative results (tab. 5) together with a remarkable proportion of false positive results. We cannot explain whether this observation is due to the selected heating of the respective test field or to a lowered cutoff value (leukocytes from 25/ μ l to 20/ μ l). Nevertheless a variation of cutoff values enables a kind of "calibration" of the instruments. Provided that a suitable calibration material is available, differences in the sensitivity of the teststrips and/or instruments could be corrected in this way. The need for a means of calibrating the instruments is also shown by the observation of discrepancies between the results of the first and the second trail. At least in part we attribute this observation to the fact that two instruments had to be exchanged during our study. We believe that our results would have been more comparable if we had been able to reliably calibrate and monitor the instruments. Mechanized teststrip readers will provide objective and reproducible urine prescreening analyses, only when these requirements are met.

These instruments measure reflectance in a continuous mode. Therefore we believe that the problems caused by the crude classification of the results (see above) can be avoided. If smaller subclasses or even a continuous mode is used for result documentation, a more quantitative analysis should be possible, and mechanized teststrip reading could show its full advantage.

A similar demand has already been made by *Haeckel et al.* in 1985 (12). Under this prerequisite, at least with respect to the concentration of protein, glucose and the liver parameters, the teststrip analysis might often replace quantitative measurements. If, for example, glucose results were given in steps of 5 mmol/l with acceptable precision, we believe that the information given to the physician would be sufficient for many clinical purposes.

With respect to the speed of analysis, we have shown that mechanized teststrip reading cannot improve the teststrip examination. Mechanized teststrip reading might show some advantages for the analysis of large series of urines. But it should be taken into account that pure teststrip reading represents only a small proportion of the overall work load in the urine laboratory. There are other tasks, like sample identification and result documentation, which — depending on individual laboratory organization — take

much more time. Therefore a common assessment of potential advantages of mechanized teststrip reading with respect to the speed of analysis is impossible. Nevertheless it is conceivable that integration of such instruments into a computerized laboratory information system would offer certain advantages versus visual reading, e.g. online sample (patient) identification, transfer and documentation of results might increase the throughput of analyses.

In conclusion we did not establish a general advantage of any one of the evaluated instruments. With respect to the validity of the results and the speed of analysis, we found no large differences between mechanized and visual teststrip reading. However if incorporated into a suitable laboratory organization the instruments might decrease the urine laboratory work load. Development of a more quantitative teststrip analysis system would realize the full potential of mechanized teststrip reading.

References

1. Hagemann, P. & Arenz, F. (1983) Teststreifen und Sedimentuntersuchung in der Urinanalyse. *Lab. Med.* 7, 39–41.
2. Bonard, C., Weber, E., Koller, P. U., Willamowski, K.-D. & Bachmann, F. (1982) Rationalisierung im Urinlaboratorium ohne Verzicht auf diagnostische Sicherheit. *Dtsch. Med. Wochenschr.* 107, 249–251.
3. Kutter, D. (1980) Teststreifen zur Rationalisierung der mikroskopischen Harnuntersuchung. *Dtsch. Med. Wochenschr.* 105, 1246–1249.
4. Kutter, D., Figueiredo, G. & Klemmer, L. (1987) Chemical Detection of Leukocytes in Urine by Means of a New Multiple Test Strip. *J. Clin. Chem. Clin. Biochem.* 25, 91–94.
5. Schaller, G. (1983) Test-Strip Screen for the Optimization of Microscopic Urinalysis. *Clin. Chem.* 29, 1692–1693.
6. Shaw, S. T., Poon, S. Y., Wang, E. T. (1985) "Routine Urinalysis" Is the Dipstick Enough? *J. Ann. Med. Ars* 253, 1596–1600.
7. Lum, G. & Morrison, M. C. (1987) Evaluation of Dipstick Urine Studies Using a Scoring System Based on Test Performance Characteristics. *Am. J. Clin. Pathol.* 84, 498–502.
8. Colombo, J. P., Peheim, E., Keller, H., Bostjancic, W., Woschnagg, B., Siest, G., Henny, J., Kaehler, K., Weiland, J., Oette, K., Schindler, J. & Wisser, H. (1983) Erfahrungen mit Kombinationsteststreifen visuell und reflektometrisch im Vergleich zum Urin-Sediment. *Lab. Med.* 7, 184–188.
9. Franzini, C., Kaltwasser, F., Leinberger, R., Nagel, D., Rodrian, J. & Seiler, D. (1985) Leistungsfähigkeit der reflexionsphotometrischen Harnanalytik. Ergebnisse der klinischen Erprobung mit Urotron® RL9. *Lab. Med.* 9, 379–383.
10. Kutter, D., Braun, C., Gallego, F. & Stirn-Thoma, S. (1988) Screening Urine before Microscopy, by Automated Test-Strip Preselection: Clinical Evaluation of the Improved Rapimat® II/T (Behring). *Clin. Chem.* 34, 1600–1602.
11. Bank, C. M., Codrington, J. F., van Diejen-Visser & Brombacher, P. J. (1987) Screening Urine Specimen Populations for Normality Using Different Dipsticks: Evaluation of Parameters Influencing Sensitivity and Specificity. *J. Clin. Chem. Clin. Biochem.* 25, 299–307.
12. Haeckel, R., Bonini, P., Ceriotti, G., Kutter, D. & Vonderschmitt, P. J. (1985) Multi-Centre Evaluation of the Urine Test Strip Analyzer Rapimat. *J. Clin. Chem. Clin. Biochem.* 23, 473–492.
13. Rattenbury, J. M. & Allen, J. C. (1987) Sample Contamination by Test Strips. *Clin. Chem.* 33, 414.
14. Heintz, R. & Althof, S. (1976) *Das Harnsediment*, 3rd edn., pp. 5–6, G. Thieme, Stuttgart.
15. SAS Institute Inc. SAS/STAT™ Guide for Personal Computers, Version 6 Edition. Cary, NC: SAS Institute Inc., 1987. 1028 pp.

Dipl.-Biol. A. M. Marx
Abteilung Klinische Chemie
und Zentrallaboratorium
Philipps Universität Marburg
Baldingerstraße
D-3550 Marburg

